

**Procédé de codage de la parole et terminaux
pour la mise en oeuvre du procédé.**

La transmission de la parole sur le réseau téléphonique commuté RTC
5 nécessite de disposer d'une bande-passante suffisante pour que la parole
reste compréhensible. Une bande allant des très basses fréquences à
quelque kilohertz représente un bon compromis entre fidélité de
restitution et ressources en bande-passante. De ce fait, pour transmettre la
voix dans le réseau RTC, dont les liaisons inter-centraux sont
10 numériques, on code les fréquences vocales pour les transformer en un
signal numérique au débit de base du réseau RTC, représentant
l'évolution de l'amplitude du signal vocal.

On souhaite cependant parfois réduire le débit de la transmission par
15 exemple dans un terminal de synthèse vocale, dont la mémoire de
messages doit rester de taille limitée. De même, on peut souhaiter
n'utiliser qu'une voie de transmission d'un réseau spécifique, à débit bien
inférieur au débit ci-dessus du réseau téléphonique.

En pareil cas, on tente de reconnaître à la volée les divers phonèmes
20 d'une séquence vocale. On compare ces phonèmes à des références, d'une
bibliothèque, associées à des mots de code et on remplace ces phonèmes
par les mots de code correspondants, qui décrivent la parole avec une
quantité d'informations bien moindre. C'est ainsi que la voix est
25 comprimée.

En réception, le terminal appelé comporte la même bibliothèque et
reconstitue, par synthèse vocale, des signaux analogiques correspondant
aux divers mots de code.

30 Mais une telle façon de procéder présente l'inconvénient de ne restituer
qu'une voix normalisée par la bibliothèque, donc impersonnelle, et il est
donc par essence impossible de reconnaître le correspondant pour
authentifier un message vocal. Les inflexions ou fluctuations de voix, qui
35 sont aussi de l'information tout comme la signification des paroles elles-
mêmes, ne sont donc pas restituées.

La présente invention vise à obtenir un codage de la voix qui permette à la fois une compression de l'information et une restitution personnalisée.

5 A cet effet, l'invention concerne tout d'abord un procédé de codage de la parole constituée d'une séquence d'unités phoniques, dans lequel on compare les unités à des références en bibliothèque associées à des mots de code primaires, on détermine les différences entre les unités et les références, on code les différences par des mots de code secondaires et on
10 substitue aux unités des paires de codes primaire et secondaire.

Ainsi, les mots de code primaires vont coder efficacement, de façon compacte, la plus grande partie de l'énergie phonique saisie tandis que les mots de code secondaires vont améliorer la fidélité de restitution sans
15 cependant accroître exagérément le volume de données de code puisqu'ils ne portent que sur une énergie limitée et qu'un faible nombre de bits permet de coder cette énergie marginale modulant l'énergie primaire, standard, correspondant aux mots de code primaires.

20 L'invention concerne aussi un terminal de codage de signaux de parole, comportant des moyens pour saisir une séquence d'unités phoniques et la transmettre à des moyens comparateurs agencés pour successivement comparer les unités phoniques à des références en bibliothèque et ainsi y sélectionner à chaque fois un mot de code primaire spécifique de l'une des
25 références, terminal caractérisé par le fait que les moyens comparateurs sont agencés pour déterminer une différence entre l'unité phonique saisie considérée et la référence correspondant au mot de code sélectionné et pour transmettre cette différence à des moyens de transcodage, prévus pour fournir en réponse un mot de code secondaire correspondant à des
30 moyens mémoires agencés pour associer les mots de code, respectivement primaire et secondaire.

L'invention concerne enfin un terminal de décodage de signaux comportant des moyens de réception de signaux représentant des mots de
35 code primaires de références d'unités phoniques en bibliothèque, et des moyens de décodage agencés pour sélectionner certaines des références

en bibliothèque d'après les mots de code primaires reçus et pour commander en conséquence un transducteur de restitution des signaux de parole, terminal caractérisé par le fait que les moyens de décodage sont agencés pour décoder en outre des mots de code secondaires, de correction, associés aux mots de code primaires, et pour corriger en conséquence les références vocales sélectionnées.

Bien que le procédé de l'invention nécessite de disposer au total d'un terminal de codage et d'un terminal de décodage correspondant, chacun de ceux-ci peut être mis séparément sur le marché et la demanderesse entend donc revendiquer l'un et l'autre.

En particulier, il est avantageux de prévoir un télécopieur comportant des moyens pour insérer les mots de code dans un message de télécopie.

L'invention sera mieux comprise à l'aide de la description suivante d'un mode de mise en oeuvre préféré du procédé de l'invention, en référence au dessin annexé, sur lequel :

- la figure 1 représente schématiquement un terminal émetteur et un terminal récepteur de signaux vocaux pour la mise en oeuvre du procédé de l'invention,
- la figure 2 représente, en fonction du temps t , l'amplitude A d'un signal de parole,
- la figure 3 représente, en fonction de la fréquence F , l'amplitude K de raies du spectre du signal de la parole 2, et
- la figure 4 est un organigramme illustrant les étapes du procédé.

Le terminal émetteur référencé 15 sur la figure 1, qui se présente ici sous la forme d'un combiné portable pour réseau de communication radio, comporte un microphone 26 de saisie du signal de parole de son utilisateur, alimentant un convertisseur analogique/numérique 27 relié en sortie à une unité centrale à microprocesseur 28 associée à deux bibliothèques 11 et 12 de séquences sonores ou unités phoniques normalisées, telles que phonèmes. L'unité centrale 28, qui code la parole, commande un émetteur 29, ici radio, dont les émissions sont reçues par

un circuit récepteur 30 d'un terminal 35 de restitution de la parole. La figure 2 représente l'amplitude d'une unité phonique en fonction du temps t et la figure 3 représente l'amplitude K des raies du spectre correspondant à un instant donné.

5

Plus précisément, l'unité centrale 28 comporte un comparateur 16 pour comparer les unités phoniques reçues du convertisseur 27 aux unités phoniques en bibliothèque. Comme expliqué plus en détails en regard de la figure 4, le comparateur 16 a pour fonction de sélectionner la référence en bibliothèque qui est la plus semblable au signal alors analysé et a aussi pour fonction de spécifier cette différence, c'est-à-dire de fournir une valeur d'écart pour chacun des critères ayant servi à la sélection. Cette différence est ici transcodée par un circuit de transcodage 17, pour condenser son expression, sous la forme d'un mot de code secondaire qui est mémorisé dans une mémoire 18 sous la commande du comparateur 16. Ce dernier, qui a préalablement mémorisé le mot de code primaire dans la mémoire 18, adresse et commande l'écriture dans celle-ci pour que les deux mots de code, primaire et secondaire, soient associés physiquement comme ils le sont d'un point de vue logique, c'est-à-dire qu'il soit par exemple défini un chaînage entre les deux zones mémoires les contenant.

Le terminal récepteur 35 comporte une unité centrale 33 effectuant un décodage inverse de la parole pour alimenter un haut-parleur 34. Deux mémoires formant bibliothèques, 31 et 32, ici externes à l'unité centrale 33, sont reliées à celle-ci. Le terminal récepteur 35 est ici un terminal classique pour la réception de messages écrits, appelé pager, agencé en outre pour recevoir des messages vocaux. Divers circuits classiques de saisie de numérotation, d'émission de celle-ci et de visualisation de données n'ont pas été représentés, dans un but de clarté.

30

L'unité centrale 33 comporte un circuit 36 d'adressage des bibliothèques 31 et 32, respectivement personnalisée et primaire, à partir des mots de code reçus du circuit récepteur 30. Un circuit tampon 37 reçoit en réponse, de la bibliothèque primaire 32, des spectres d'unités phoniques primaires et les transmet à un circuit 38 de modulation ou composition de

35

ces spectres. Le circuit 38 module ceux-ci en fonction du mot de code secondaire associé au mot de code primaire de lecture de la bibliothèque primaire 32. Le circuit 38 combine ainsi les informations des mots de code primaires et secondaires pour reconstituer le signal de parole capté à l'origine (26). Cette combinaison peut par exemple être une addition ou une multiplication, de raies de fréquences suivie d'une tranformation de Fourier inverse ou, encore, elle peut porter directement sur des amplitudes de signal. Dans cet exemple, chaque type d'unité phonique restituée est mémorisée dans la mémoire personnalisée 31, afin d'utiliser directement cette dernière si un même couple de mots de code, primaires et secondaires, est ultérieurement reçu. En variante, la mémoire 31 pourrait ne contenir que des valeurs de modulation, qu'elle fournirait au circuit 38 après adressage par un mot de code secondaire.

Les opérations de codage et de décodage va maintenant être exposées plus en détails, en référence à la figure 4.

Pour coder la voix, on capte à une étape 1, un signal de parole par le microphone 26 et, ici, on le convertit en un signal numérique dans le convertisseur 27, à une étape 2. Le signal de parole est alors comparé, dans l'unité centrale 28, à une pluralité de signaux de référence de la bibliothèque 11, à une étape 3. La comparaison intervient à la volée, en pratique cycliquement à vitesse élevée par rapport à la vitesse d'évolution du signal de parole analysé. Celui-ci peut être considéré comme étant une séquence d'unités phoniques, spécifiques d'une langue donnée, telles que voyelles, diphtongues ou hiatus, dont une représentation a été initialement mise en bibliothèque 11 et associée à un mot de code, dit primaire, particulier à chacune. Lors de la constitution de la bibliothèque 11 et des bibliothèques 12 et 32 indiquées plus loin, on effectue plusieurs saisies vocales d'un même locuteur ou de plusieurs pour constituer une référence vocale moyenne. De préférence cependant, afin d'améliorer l'efficacité de la reconnaissance future, on mémorise (11, 12), pour chaque unité phonique, plusieurs références pour constituer un domaine de reconnaissance permettant de tolérer des écarts entre locuteurs.

Chaque unité phonique (fig. 2) correspond à une évolution particulière de l'amplitude A ou énergie du signal de parole et présente une durée susceptible de varier selon la vitesse d'élocution de la personne qui parle.

5 L'étape 3 consiste donc à comparer l'évolution des amplitudes des signaux de référence à celle du signal capté. Pour s'affranchir des variations de vitesse d'élocution, on peut par exemple ne considérer que la succession des modulations d'amplitude significatives (variation d'énergie dépassant un seuil), sans y associer une notion de valeur de
10 temps.

Sur la figure 2, les flèches verticales, ici au nombre de huit, représentent l'amplitude des extréma et forment ainsi une signature, supposée ici représenter une unité phonique déterminée.

15 Si, quittant le domaine temporel de la figure 2, on considère maintenant le domaine fréquentiel, la transformée de Fourier de l'amplitude instantanée A du signal en un point quelconque de la courbe de la figure 2 peut être représentée par le spectre de raies de fréquences de la figure 3. En
20 pratique, on considère que l'énergie vocale se trouve essentiellement limitée à trois bandes de fréquences, respectivement situées vers 0,1 kHz et deux bandes entre environ 1 et 3 kHz ainsi que 5 et 7 kHz, respectivement.

25 De ce fait, si l'on parcourt la courbe des amplitudes de la figure 2 en fonction du temps t , l'amplitude K de chaque raie de la figure 3 va se trouver modulée en fonction de l'évolution de l'amplitude A du signal de parole.

30 Ainsi, si l'on mémorise la succession des spectres de la figure 3, on peut reconstituer la succession des amplitudes A du signal d'origine, par transformée de Fourier inverse.

35 Afin de limiter le nombre de spectres à traiter, on peut n'effectuer que des échantillonnages cycliques, suffisamment rapprochés pour ne pas perdre d'information. On peut encore se limiter à un jeu restreint de

spectres des extréma d'amplitude représentés par les huit flèches de la figure 2. Si l'on souhaite encore limiter le nombre de spectres, on peut même ne retenir qu'un seul spectre représentant la moyenne de tous les spectres sur la période de temps considérée de l'unité phonique ou la
 5 moyenne des spectres des extréma.

Comme cela est expliqué plus loin en détail, on va comparer le spectre moyen, ou les spectres, du signal de parole capté à un ou des spectres homologues de signaux de parole de référence en bibliothèque pour,
 10 d'une part, sélectionner le signal de parole de référence (unité phonique) le plus semblable au signal capté et, d'autre part, élaborer un signal de différence entre le ou les spectres de ce dernier et le ou les spectres du signal de référence sélectionné. Le signal de différence est mis sous
 15 forme de mot de code, dit secondaire, et est associé au mot de code primaire du signal de référence sélectionné (unité phonique reconnue) et constitue ainsi une information additionnelle de modulation ou correction du signal analogique normalisé qui sera restitué à partir du mot de code primaire considéré.

20 Les mots de code primaires des unités phoniques successivement sélectionnés, à mesure que se déroule la séquence vocale, sont mémorisés à une étape 4 pour constituer un message, codé selon la norme de la bibliothèque 11.

25 Par ailleurs, à une étape 5, on traite plus avant certaines des unités phoniques captées et reconnues, en analysant en détail, ici dans le domaine fréquentiel par transformée de Fourier inverse, leur spectre en fréquence, comme expliqué plus haut, étape 6.

30 A une étape 7, on compare le spectre de raies j de l'unité phonique d'identité i concernée, ou les spectres représentant son évolution au cours du temps t , au(x) spectre(s) de l'unité phonique sélectionnée en bibliothèque 11, contenu(s) dans la bibliothèque associée 12. On établit ainsi pour le, ou chaque, spectre, une série de coefficients de pondération
 35 Cijt (i = identité du phonème, j = rang fréquentiel de la raie, t = rang temporel) indiquant chacun l'amplitude ou énergie relative de chaque raie

j par rapport à son homologue en bibliothèque 12. Exprimé autrement, ces coefficients représentent aussi, bien qu'indirectement, la différence relative ($1 - C_{ijt}$) entre l'unité phonique reconnue et la référence correspondante en bibliothèque). Les raies dans chacune des trois bandes
 5 correspondent en fait à une rangée de mini-bandes de fréquences adjacentes, dans lesquelles on détecte de l'énergie vocale. L'analyse dans le domaine fréquentiel, ici choisie, fournit ainsi une information plus détaillée que dans le cas d'une analyse dans le domaine temporel de la figure 2, où l'on ne dispose que de l'amplitude instantanée A.

10 Ainsi, dans le cas des figures 2 et 3, la série ci-dessus comporte douze coefficients représentant les douze raies dessinées, si bien qu'un tableau de huit telles séries représente l'unité phonique, à travers les huit extréma dessinés. Hormis la réduction du tableau, à une seule série, on peut
 15 prévoir de ne retenir qu'un coefficient de pondération moyen pour chacune des trois bandes. Si l'on code chaque coefficient sur simplement 4 bits, l'erreur ne dépasse pas 3 % environ, ce qui est amplement suffisant pour restituer un timbre de voix, d'autant que le signal de correction représente peu d'énergie par rapport au signal normé qu'il
 20 corrige, si bien que l'erreur, rapportée au total, est faible.

On peut ainsi associer ici, au mot de code primaire de l'unité phonique sélectionnée, de l'ordre de la centaine de bits (12×8) si l'on retient chaque extrémum, ou seulement 12 bits (4×3) pour les trois bandes.
 25 Comme le timbre de la voix est surtout fourni par les fréquences hautes de la troisième bande, on peut même ne transmettre que le mot de code secondaire, de correction, relatif à celle-ci.

A une étape 8, on transforme le signal de différence de spectres en un
 30 mot de code secondaire, représentant le tableau ou la série ci-dessus. Lorsque la séquence de parole captée s'achève, les mots de code primaires de l'étape 4 et les mots de code secondaires de l'étape 8 sont associés un à un (étape 9) puis émis sur un réseau de transmission, comme par exemple le réseau téléphonique commuté ou, ici, un réseau de
 35 radio messagerie (étape 10).

Le terminal appelé 35 reçoit le message à une étape 21 et, à une étape 22, un fichier en bibliothèque primaire 32, semblable au fichier de spectres 12, est lu par le circuit 36 pour en extraire les spectres normalisés, primaires, d'après les mots de code primaires. A une étape 23, les mots
5 de code secondaires servent à moduler (38) les amplitudes ou énergies des raies normalisées lues en bibliothèque primaire 32, pour ainsi constituer la bibliothèque personnalisée, 31, d'unités phoniques, c'est-à-dire comportant en particulier le timbre de la voix captée. Les unités
10 phoniques de la bibliothèque personnalisée 31 sont représentées sous forme numérique dans le domaine temporel, après une transformation préalable par transformée de Fourier inverse.

A une étape 24, les mots de code primaires reçus sont lus successivement pour restituer, par le haut-parleur 34 (étape 25), le signal de parole capté.
15 Pour cela les mots de code primaires lisent la bibliothèque personnalisée 31, qui correspond donc à la bibliothèque 11 mais qui a été personnalisée par les caractéristiques en spectre de la voix captée.

La constitution de la bibliothèque 31 est, comme évoqué plus haut, facultative et a pour but de mémoriser une correction pour chaque mot de
20 code primaire, ce qui évite de répéter l'envoi du mot de code secondaire lorsqu'un même mot de code primaire est émis plusieurs fois. Si, au contraire, on émet systématiquement un mot de code secondaire, celui-ci peut évoluer pour suivre les éventuelles évolutions du timbre. Dans ce
25 cas, la voix restituée est à la fois personnalisée et en outre on restitue l'évolution du timbre au fil du temps.

Il faut en outre remarquer que, de façon générale, l'analyse et la restitution peuvent porter sur toute la bande de fréquences audibles, allant
30 d'environ 15 Hz à 15 kHz, même si en pratique on peut se limiter à 8 kHz. Les fréquences de la bande allant de 4 à 8 kHz, coupées pour les transmissions classiques par le réseau téléphonique, sont ici analysées et restituées puisqu'on transmet l'information correspondante sous la forme d'une télécommande de la bibliothèque 31, qui contient déjà les raies à
35 ces fréquences, ce qui évite toute transmission explicite de celles-ci.

Il faut par ailleurs remarquer que, si l'analyse peut ne porter que sur un nombre limité de bandes de fréquences suffisamment caractéristiques en bibliothèque 11, 12, les divers signaux à restituer, en bibliothèque 32, contiennent la totalité des raies initialement saisies, c'est-à-dire couvrent chacun par exemple une bande d'un seul tenant de 15 Hz à 8 kHz.

Comme exposé au début, l'invention peut s'appliquer en dehors de toute transmission, pour par exemple mémoriser localement un message à restituer ultérieurement, c'est-à-dire qu'il s'agit alors de la fonction d'un magnétophone.

Dans une autre forme de réalisation, non représentée, les mots de code, primaire et secondaire, sont associés à des données de télécopie pour former un message voix-données. Le message est saisi par le poste téléphonique usuellement associé aux télécopieurs et est restitué par le même moyen au niveau du télécopieur appelé. Les mots de code, émis par un circuit comme 28, sont insérés dans un champ spécifique du message par un microprocesseur gérant le protocole de transmission de télécopies et sont de même extraits en réception pour être traités comme expliqué plus haut. On peut ainsi réaliser une annotation vocale d'un message de télécopie, annotation transmise par exemple comme un en-tête de télécopie.

REVENDICATIONS

1.- Procédé de codage de la parole constituée d'une séquence d'unités phoniques, dans lequel on compare les unités à des références en bibliothèque associées à des mots de code primaires, on détermine les différences entre les unités et les références, on code les différences par des mots de code secondaires et on substitue aux unités des paires de codes primaire et secondaire.

2.- Procédé selon la revendication 1, dans lequel, la comparaison portant sur les énergies de spectres de raies de fréquences, on détermine, pour les raies, des coefficients de pondération, normés par rapport à l'énergie des raies de référence, et on intègre lesdits coefficients dans le mot de code secondaire.

3.- Procédé selon l'une des revendications 1 et 2, dans lequel on détermine ladite différence à partir d'une succession de spectres correspondant à une succession d'amplitudes de l'unité phonique considérée.

4.- Procédé selon la revendication 3, dans lequel on ne considère que les amplitudes correspondant à des extrema.

5.- Procédé selon la revendication 2, dans lequel on détermine ladite différence à partir d'un spectre unique moyen de l'unité phonique considérée.

6.- Procédé selon l'une des revendications 2 à 5, dans lequel on limite la comparaison fréquentielle à trois bandes de fréquences.

7.- Procédé selon la revendication 6, dans lequel on exprime le coefficient de pondération des raies de chaque bande par un unique coefficient.

8.- Terminal de codage de signaux de parole, comportant des moyens (26, 27) pour saisir une séquence d'unités phoniques et la transmettre à des moyens comparateurs (16) agencés pour successivement comparer les

unités phoniques à des références en bibliothèque (11, 12), et ainsi y sélectionner à chaque fois un mot de code primaire spécifique de l'une des références, terminal caractérisé par le fait que les moyens comparateurs (16) sont agencés pour déterminer une différence entre l'unité phonique saisie considérée et la référence correspondant au mot de code sélectionné et pour transmettre cette différence à des moyens de transcodage (17), prévus pour fournir en réponse un mot de code secondaire correspondant à des moyens mémoires (18) agencés pour associer les mots de code, respectivement primaire et secondaire.

9.- Terminal de décodage de signaux de parole, comportant des moyens (30) de réception de signaux représentant des mots de code primaires de références d'unités phoniques en bibliothèque (32), et des moyens de décodage (33) agencés pour sélectionner certaines des références en bibliothèque (32) d'après les mots de code primaires reçus et pour commander en conséquence un transducteur (34) de restitution des signaux de parole, terminal caractérisé par le fait que les moyens de décodage (33) sont agencés pour décoder en outre des mots de code secondaires, de correction, associés aux mots de code primaires, et pour corriger en conséquence (38) les références vocales sélectionnées.

10.- Télécopieur selon l'une des revendications 8 et 9, comportant des moyens pour insérer les mots de code dans un message de télécopie.

ABREGE

Procédé de codage de la parole et terminaux pour la mise en oeuvre du procédé.

5

Le procédé de codage de la parole, constituée d'une séquence d'unités phoniques, consiste à comparer les unités à des références en bibliothèque associées à des mots de code primaires, à déterminer les différences entre les unités et les références, on code les différences par des mots de code secondaires pour substituer aux unités des paires de codes primaire et

10 secondaires et, si la comparaison porte sur les énergies de spectres de raies de fréquences, on détermine, pour les raies, des coefficients de pondération, normés par rapport à l'énergie des raies de référence, et on intègre les coefficients dans le mot de code secondaire.

15

Figure 1